DOCUMENT RESUME

ED 165 837 JC 790 111

AUTHOR Baratta, Mary Kathryne
TITLE Magnitude Discrimination Index: A Merger of Delphi Assumptions and Realities.
INSTITUTION Moraine Valley Community Coll., Oak Lawn, Ill.
PUB DATE 79
NOTE 15p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Decision Making; *Discriminant Analysis; Factor Analysis; *Feedback; *Management; Management Games; *Mathematical Models; Multidimensional Scaling; Problem Solving
IDENTIFIERS *Delphi Technique

ABSTRACT

The advantages of the Delphi Technique are that it relies on collective expert judgment, prevents the excessive influence in decision-making of those in positions of status, and encourages consensus. The technique consists of five steps: (1) asking experts to list items relevant to a topic; (2) returning the list to the participants and asking them to evaluate it according to given criteria; (3) returning the list, with a summary of responses, to the participants and asking those with minority responses to revise or defend their position; (4) returning the list and an updated summary of majority and minority responses again to provide additional opportunity for revision; and (5) formalizing a report of results arrived at by the repeated questionings. Traditionally, three major conventions are used in the final reporting: reporting by importance, reporting by consensus, and reporting by both consensus and importance. An additional alternative, which combines the elements of consensus and importance in one measurement, is a magnitude discrimination index. The advantages of this method of reporting are explored in relation to the other three methods. A bibliography is included. (MB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED165837

Magnitude Discrimination Index: A Merger of Delphi Assumptions and Realities

Mary Kathryne Baratta, Ph.D.
Moraine Valley Community College
10900 South 88th Avenue
Palos Hills, Illinois 60465

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Mary Kathryne Baretta

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

JC 790 111

1979

Abstract

Since its inception at The Rand Corporation, the Delphi Technique has been used to gather and report the informed opinions of experts. The Delphi Technique is based upon three assumptions: (1) it relies on collective expert judgement; (2) it prevents professional status and high position from forcing judgements in certain directions and (3) it implies that groups with disparate judgements move toward consensus.

Although the three assumptions upon which the Delphi Technique is based remain constant, there are variations in both the methodology used and the convention of reporting final results. A review of the major reporting conventions and an alternative method for reporting Delphi results is the major focus of this paper.

Magnitude Discrimination Index:
A Merger of Delphi Assumptions and Realities

Since its inception at the Rand Corporation (Helmer, 1966), the Delphi Technique has been used to gather and report the informed opinions of experts. The Technique involves: (1) asking several experts to list items they feel are relevant to a specific topic, (2) returning to the participants the list of items generated and asking them to evaluate the total list according to a set of criteria; (3) returning to the participants the list of items again with a summary of responses to the items and, if in the minority, the participants are asked to revise their minority opinions or indicate their reasons for not choosing the majority opinion; (4) returning to the participants again the list of items, an updated summary and minority opinions thus providing another chance for the participants to revise their opinions. Step four may be repeated several times and (5) formalizing in report form the results arrived at by the repeated questioning.

The Delphi Technique is based upon three assumptions: (1) it relies on collective expert judgement; (2) it prevents professional status and high position from forcing judgements in certain directions, as frequently occurs when panels of experts meet and (3) it implies that groups with disparate judgements move toward consensus.

Although the three assumptions upon which the Delphi Technique is based remain constant, there are variations in both the methodology used and the convention of reporting final results. Methodological changes involve either elimination of the first step, generation of a list of items or a change in the total number of rounds used. The reporting of final results can be categorized into three major conventions: reporting by importance, reporting by consensus and reporting by importance and consensus.

A review of the major reporting conventions and an alternative method for reporting Delphi results is the major focus of this article.

Typical Conventions of Reporting Delphi Results

Reporting by Importance

The accepted practices of reporting final Delphi results by importance include the use of mean, median and mode. These measures of central value refer to the location of the center of a distribution of responses.

The mean was used to report Delphi results by the Educational Research Corporation to evaluate alternative schools (Willard and Glick, 1975). The mean of each item was obtained by summing all of the observations for an item and dividing by the total number of observations. These means were then rank ordered. This ranking procedure involves assigning to the item with the largest mean value a rank of one (assuming an ascending rating scale of importance), assigning the item with the next largest mean value a rank of two, etc. After ranking the means the Delphi results are analyzed. A rank of one denotes the item of greatest importance to the participants in the study.

The ranking of medians is also used to report Delphi results. This method was used by The Ohio State University to help an intermediate school district with planning (Anderson, 1970). The median is the middle observation if there is an odd number of cases and, by statistical convention, is the mean of two central observations if there is an even number of cases. The medians for each item are then rank ordered according to importance, i.e., the larger the value of the median the greater the importance.

Another accepted practice for reporting the importance ratings of Delphi results is the mode. This method was used by Atlanta Assessment Project to establish educational goals (Sweigert Jr. and Schabacker, 1974). By definition, the mode is the score with the greatest frequency. The importance ratings are obtained by rank ordering the modes for the various items in decreasing order.

Depending on the nature of the data and on its uses, some measures of central tendency are more appropriate under certain conditions than others. The mode tells us nothing more than the category or item which contains the largest number of cases. For example, if the distribution for an item on a Delphi study was 7 A's, 9 B's, 12 C's, 8 D's and 4 E's, the choice "C" would be the mode of the distribution for the item since it occurred most often. The mode is not affected by extreme scores or by any value other than the frequency of the scores. It is not necessary to know the extreme scores or the distances of the scores from one another to determine the mode. When the number of cases in a distribution is small, or in situations calling for fine grouping, the mode may have little meaning. The concept of mode as a measure of central tendency is chiefly useful because it invites attention to bimodal data.

The median is a point on the scale which divides the distribution of the data into two equal parts. One half of the cases fall above the median and one half of the cases fall below the median. When a distribution contains a few extremely low or extremely high scores as compared to the rest of the data, the median provides a better representation of the data because it is not sensitive to the extreme scores. Another instance when the median is appropriate is in an open-ended distribution (a distribution with unspecified top or bottom).

The mean is the most widely used and useful measure of central tendency. It is the most reliable measure, as a rule, and is simply and clearly defined. However, when a series include a few items of either high or low values relative to the values of the majority, the mean is not a fair measure of central tendency.

Ranking by Consensus

When a researcher is confronted with the problem of data reduction and description, measures of dispersion or variability are used. A measure of dispersion reflects the extent to which the measurements or scores differ from each other. Two types of variability commonly used with the Delphi Technique are the interquartile range and the standard deviation. These measures are used when discussing consensus, the degree or amount or agreement.

The interquartile range is bounded by the seventy-fifth percentile and the twenty-fifth percentile, i.e., it includes the group of scores between these values. With successive steps of the Delphi this interval should converge, the shrinking of the interquartile range may be defined as consensus. The various interquartile ranges are then ranked; the smaller the interval the larger the degree of consensus or agreement. The interquartile range is meaningful only for data that are measured on at least an ordinal scale and is commonly used when the median is the location of interest. It excludes extreme scores and is a function of only two points on a scale.

Another method of describing dispersion is the standard deviation. The standard deviation is the square root of the average squared deviation from the mean. The standard deviation is reported as a number, the smaller the value of the number the greater the consensus. The standard deviations are then ranked from the smallest value, implying a higher degree of agreement,

to the largest value implying a lesser degree of agreement. On successive rounds of the Delphi the standard deviation of any item should be reduced in value to indicate increasing consensus. The standard deviation provides a one-dimensional measure of dispersion and is commonly used with the mean.

Combination of Importance and Consensus

The concept of consensus reflects the extent of agreement among respondents about importance, but it is independent of the level of importance. Consensus was developed as a spin off of the work with the Delphi Technique and has been included as part of the reporting of results concerning Delphi studies. Although many Delphi studies report only the mean, median or mode, others report median and interquartile range or mean and standard deviation. When reporting the interquartile range in addition to the median or the standard deviation in addition to the mean, both importance and consensus are examined.

The Rand Corporation (Helmer, 1967) used a combination of median score and interquartile range to analyze the future of automation. The median and interquartile range were fed back to respondents during each round. The median of the final group response was taken as the nearest thing to group consensus. The median and interquartile range were then reported for each item in the final report.

Gross and Grambsch (1968), Uhl (1971) and Harshman (1974) used the mean as a measure of importance or priority and the standard deviation as a measure of consensus. The final report consisted of items ranked by mean importance; the standard deviation about each mean was also listed.

The above methods of reporting Delphi results supply the user with both importance and consensus information. The reduction of a distribution to a measure of central tendency is not without danger of distorting information.

Variability is an important feature of a distribution. A mean conceals this feature, and a comparison of mean values may be unfair and misleading if the distributions are dissimilar in variability. A measure of central tendency does not have meaning independent of the other characteristics of a distribution; in fact, if a series is highly variable or irregular and rich in detail, a mean may have no real meaning and serve no useful purpose at all.

An Alternative Method of Reporting Delphi Results

Before developing another method of reporting Delphi results, two issues need to be examined. First, a method for combining the two types of information, importance and consensus, is necessary because one scale is in ascending order and the other scale is in descending order. On a scale measuring importance, the larger the numerical value the greater the importance; whereas, on a scale measuring consensus, the smaller the numerical value the greater the consensus thus causing trouble when a summation of these numerical values is used. Summing scores for items with a high importance value and a high consensus value result in the importance value being disproportionately weighted; whereas, combining the scores for items with low importance value and low consensus value result in the consensus value being disproportionately weighted.

Next, the order of magnitude of the two scales differ, i.e., importance scores have a limited range of values; whereas, consensus scores may range from zero to infinity. When summing these two values, giving equal weight to both, the consensus value may overshadow any effect of the importance value. A new ranking method has to take into account these effects.

Magnitude Discrimination Index

When examining importance and consensus, the "ideal" situation would be to have the item(s) with the highest importance rating to also have the largest amount of consensus, everyone would agree about the importance of the item. However, such is not always the case, for some items have a high importance rating with little agreement (consensus), while others have a lower importance rating with a great deal of agreement or consensus. This new ranking method takes into consideration both consensus and importance, weighting them equally.

First, importance will be measured by the mean and consensus will be measured by the convergence score. The mean was selected because it is the most reliable measure of central tendency. The convergence score was selected because it is a measure of variability that is distribution free. The convergence score is defined by Uhl (1971) as the absolute sum of the distance between each participant's response and the mean of all participants' responses across all items.

For example if item X had responses of 3, 5 and 1, the mean would be 9/3 or 3. To obtain the convergence score, the mean, 3 in the example, would be subtracted from each respondent's choice, i.e., 3-3, 5-3, 1-3. The resulting subtraction disregarding signs yields 0, 2 and 2; these values are then summed, 4, and this is the convergence score for item X.

To arrive at a final ranking using the new method the following steps must be taken:

(1) rank order the importance rating ($\bar{x}$) from most important to least important, assigning a rank of one to the item with the greatest amount of importance;

(2) rank order the consensus values, as measured by the convergence score, from greatest consensus to least consensus assigning a rank of one to the item with the greatest consensus;

(3) sum the ranks for importance and consensus for each item;

(4) re-rank the sum of ranks to arrive at the grand ranks, the basis for the final rankings.

If the sum of ranks combined the most important mean with the largest amount of consensus for each item, the sum of ranks for ten items would have the form 2,4,6,...,20, that is, 1+1=2, 2+2=4,...,$n+m=R_G$, an ideal situation. Variations in this series represent departures from the ranking agreement between the mean and convergence scores.

Ranking scales are order-preserving transformations, that are "unique up to a monotonic transformation." It does not matter what numbers are given to a pair of classes, or to members of those classes, just as long as the higher rank is given to the members of the class which are "greater" or "more preferred" (Siegal, 1956). This new method of ranking is based on this premise, thus resolving the problems of reverse scales and magnitude of scales.

The method of ranking just described gives equal weight to consensus and importance; however, this ranking method can be expanded to take into consideration unequal weights for importance and consensus. For example, the importance rating could be weighted twice as much as the consensus rating, in this case all ranks for consensus would be multiplied by two (the smaller the grand rank the better the item). Matters of unequal weighting are left to the disgression of the reader.

The Magnitude Discrimination Index is an Improvement

Delphi data has been ranked on the basis of importance ratings, consensus scores and a combination of importance ratings and consensus scores. These methods differ in their consideration of the degree of importance and the degree of consensus. Since several different methods for ranking information collected by the Delphi procedure were suggested and examined only on an intuitive level, the next level of evaluation is comparison of the various methods.

Magnitude Discrimination Index vs. Importance

Aspects of importance are merely a rank ordering of a measure of central tendency (mean, median, mode). When the mode is used to report Delphi results, problems arise if there are several modes for one item, how is this occurrence noted--no consensus? Problems also occur with the definition of majority, is it 50%, 51%, 60%? How is consensus measured? Is consensus implied?

The median as a method of reporting final Delphi results is an improvement over the mode because it eliminates the possibility of a bimodal importance rating; however, it does not account for the clustering of scores, for example the scores may cluster at two end points of the scale. Again how do we know if consensus was truely achieved?

When the mean is used to report results, a mean larger in value is always ranked ahead of a mean smaller in value. Ranking importance in terms of mean scores does not take into account the amount or degree of agreement as to the most important item, a rarely advisable decision. Statistically, when interpreting any mean value the distribution of ratings must be taken into consideration.

These methods of reporting the final results of the Delphi Technique only imply consensus but, they do not take into consideration the degree of consensus. An implication of the Delphi Technique is that it produces consensus, but in the reporting of the results, this aspect is taken for granted.

The magnitude discrimination index is an improvement because it takes into account both importance and consensus. It also eliminates the inherent problems of working with the mode and median. By reporting both the mean and the convergence score the center of the distribution of scores as well as how they are clustered are reported, thus providing more information on which to base decisions.

Magnitude Discrimination Index vs. Consensus

Two of the methods used to measure consensus were the interquartile range and the standard deviation. The interquartile range does not take into account the clustering of scores, for example, the scores may cluster at opposite ends of the scale in which case the interquartile range has little meaning. The standard deviation indicates the spread of scores from the mean; however, it assumes that the data from an item will be normally distributed. A bimodal distribution may have a mean and standard deviation which are the same as those for a normal distribution. However, a bimodal distribution indicates that the perceptions are polarized, reflecting an actual lack of consensus.

If ratings were based solely on consensus, items could have consensus scores of zero (rank 1) and may have a low importance rating, i.e., everyone agrees that it is not important. This method of ranking does not take into account importance and can result in decision-making errors. Although consensus is a necessary condition of the Delphi Technique, by itself it is not sufficient and leaves us without valuable information on which to base decisions.

Again by using the magnitude discrimination index there is an improvement in the reporting of Delphi results because it accounts for both importance and consensus thus providing the users with additional information.

Magnitude Discrimination Index vs. A Combination of Importance and Consensus

Reporting the median and interquartile range or the mean and standard deviation are improvements over reporting only importance or consensus. But the problems that existed for importance or consensus are compounded. Also, reporting two measures, importance and consensus, tends to be confusing. For instance, how would an item rank if it was high in importance and low in consensus? By using the new ranking method, a combination of mean scores and convergence scores, some of the above mentioned problems are eliminated because the convergence score is distribution free. The new method of ranking looks at importance and consensus simultaneously by ranking those items with a high degree of importance and strong consensus ahead of those items with a high degree of importance and moderate consensus, a high degree of consensus and moderate importance and finally a lesser degree of importance and very strong consensus.

Several methods of ranking Delphi Technique data have been described and compared. Importance reporting fell short of providing sufficient data on which to base decisions in and of itself; while consensus reporting lacked some necessary data on which to base decisions. Combinations of importance and consensus produced problems in interpretation. The proposed new method of ranking Delphi results by the magnitude discrimination index takes into account both importance and consensus, and has the additional quality of being distribution free.

MKB/st

References

Anderson, D.P. "Ohio State Study: Clarifying and Setting Priorities on an Intermediate School District's Objectives Utilizing the Delphi Technique." A paper delivered at the American Educational Research Association meeting March 4, 1970 in Minneapolis, Minnesota.

Baratta, M.K. Utilization of the Delphi Technique to Weight Program Evaluation Criteria in Vocational Education. (Doctoral dissertation, The Ohio State University) University of Microfilms: Ann Arbor, Michigan, 1974, No. 75-3008.

Cyphert, F.R. and Gant, W.L. "The Delphi Technique: A Case Study." Phi Delta Kappan, Vol. 52, 1971, pp. 272-273.

Emmons, J.I. and Kaplan, L. The Delphi Method and Decision-Making: A Futuristic Technique. International Communication Association: Phoenix, Arizona, 1970.

Gross, E.W. and Grambsch, P.V. University Goals and Academic Power. American Council on Education: Washington, D.C., 1968.

Helmer, O. "The Use of the Delphi Technique in Problems of Educational Innovations." The Rand Corporation: Santa Monica, California, 1966.

Helmer, O. "Analysis of the Future: The Delphi Method." The Rand Corporation: Santa Monica, California, 1967.

Hudspeth, D.R. "A Long Range Planning Tool for Education: The Focus Delphi." Syracuse University Research Institute: Syracuse, New York, 1970.

Judd, R.C. "Delphi Method: Computerized 'Oracle' Accelerates Consensus Formation." College and University Business, September 1970, Vol. 49, No. 3, pp. 30-34.

Siegal, S. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill, Inc.: New York, New York, 1956.

Sweigert, R.L. and Schabacker, W.H. "The Delphi Technique: How Well Does it Work in Setting Educational Goals." A paper presented at the American Educational Research Association meeting April 1974 in Chicago, Illinois.

Uhl, N.P. Identifying Institutional Goals. National Laboratory for Higher Education: Durham, North Carolina, 1971.

Willard, R.W. and Glick, L.J. "Alternative Uses of the Delphi Technique in Evaluating Alternative Schools." A paper presented at the American Educational Research Association meeting March 31, 1975 in Washington, D.C.